DOCUMENT RESUME

ED 337 150 IR 015 201

AUTHOR Collins, Allan; And Others
TITLE Three Different Views of Students: The Role of Technology in Assessing Student Performance. Technical Report No. 12.
INSTITUTION Center for Technology in Education, New York, NY.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Apr 91
CONTRACT 1-35562167-A1
NOTE 11p.
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Comparative Testing; *Computer Assisted Testing; Computer Simulation; Elementary Secondary Education; Microcomputers; Problem Solving; Science Tests; Test Construction; *Test Format; *Videotape Recordings
IDENTIFIERS *Paper and Pencil Tests

ABSTRACT

The use of paper and pencil, videotape recordings, and microcomputers in student testing provide three very different views of student achievement. Paper and pencil tests can record how students compose tests and documents, and how they critique documents or performances. Video recordings can record how students explain ideas, answer questions, listen in class, cooperate in a joint task, and perform experiments. Microcomputers can record how students learn with feedback, trace their thinking processes, and record their abilities to deal with realistic situations. The construction of tests and the format those tests take are based, in part, on the kinds of tasks that students are required to learn to perform and a change in educational testing affects the instructional material that is taught. A key part of a current effort to develop systemically valid methods of assessing student performance in the context of high school science is the exploration of the kinds of tasks that will enable students to use and demonstrate the broader range of abilities outlined above, which will require very different kinds of tasks than those that are now the norm. Data are being collected for this project using a computer program called Physics Explorer, which provides students with a simulation environment in which they can conduct experiments to determine how different variables affect each other in a physical system. Another kind of task might be diagnosing why a system is not behaving as expected. Computers can also provide a setting where students can carry out a design task and try it out in a simulation. Research needs to be conducted to test this new approach to testing in schools, paying particular attention to its reliability, generalizability across schools, and effectiveness. (17 references) (DB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED383150

CENTER FOR TECHNOLOGY IN EDUCATION

Three Different Views of Students: The Role of Technology in Assessing Student Performance

Allan Collins, Jan Hawkins, and John R. Frederiksen

BANK STREET COLLEGE OF EDUCATION
610 WEST 112th STREET
NEW YORK, NY 10025

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

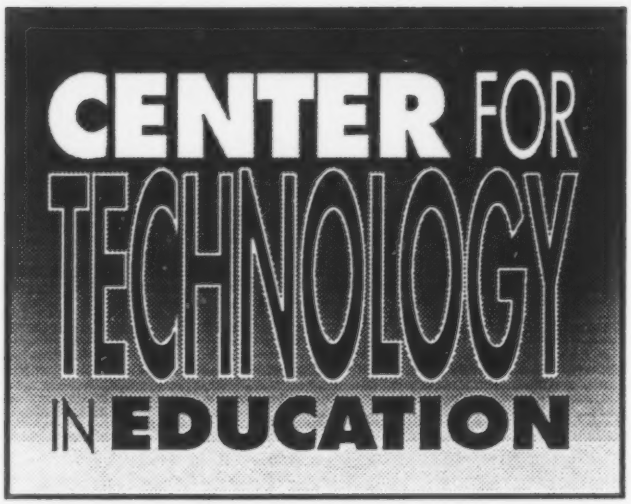
K. McMillan

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

1201T201

BEST COPY AVAILABLE

Technical
Report
Series



Technical Report No. 12

Three Different Views of Students: The Role of Technology in Assessing Student Performance

Allan Collins, Jan Hawkins,
and John R. Frederiksen

April 1991

Tech. Rep. No. 12

THREE DIFFERENT VIEWS OF STUDENTS: The Role of Technology in Assessing Student Performance

Allan Collins
Bolt Beranek and Newman Inc.

Jan Hawkins
Bank Street College of Education

John R. Frederiksen
Educational Testing Service

INTRODUCTION

If you asked scientists what qualities make a good scientist, they might come up with the following list: the ability to explain ideas and procedures in written and oral form, to formulate and test hypotheses, to work with colleagues in a productive manner, to ask penetrating questions and make helpful comments when you listen, to choose interesting problems to work on, to design good experiments, and to have a deep understanding of theories and questions in your field. Excellence in other school subjects, such as math, English, and history, requires similar abilities.

If you think about how to assess such an array of different knowledge and abilities, it is clear that paper and pencil cannot in any direct way assess most such abilities. And yet our entire testing system is almost completely reliant on paper and pencil. It is really as questionable as trying to judge a gymnast's or musician's ability with paper-and-pencil testing. Paper and pencil can only measure a small part of mathematical, scientific, and language ability.

And yet everyone agrees that tests have a large effect on what is taught. Administrators, teachers, and students will emphasize those abilities necessary to do well on tests, and the pressures to do so are becoming more intense. If the testing system only taps a small part of what it means to know and do science or math or English or history, then testing will drive the system to emphasize a small range of those abilities. We would argue that it in fact has done just that. In science, the paper-and-pencil testing system has driven education to emphasize just two abilities: recall of facts and concepts, and ability to solve short, well-defined problems. These two abilities do not, in any sense, represent the range of abilities required to be a good scientist.

We would argue that it is proper for assessment to drive the education system. People need goals as to what they should be learning, and tests encapsulate abstract learning goals in a concrete form that everyone can understand. But there is a huge disparity between the goals realized in the current paper-and-pencil tests and the authentic goals of education we

This work was supported by the Center for Technology in Education under Grant No. 1-35562167-A1 from the Office of Educational Research and Improvement, U.S. Department of Education, to Bank Street College of Education.

should be pursuing as a society: to teach people how to learn and think like scientists, writers, bookkeepers, technicians, etc. In our view, education should pursue the goals of being a thoughtful citizen who can meet the changing demands of society (Collins, in press; Zuboff, 1988).

Our thesis is that paper and pencil, video, and computers give three very different views of what students can do. It is like three different camera angles on the complete picture of a student. Whereas you cannot possibly reconstruct the total person from just one angle, with three different views you can triangulate to get a much richer notion of what a student's abilities are. By enriching the way we assess students, we will enrich the way we educate them.

STORIES ABOUT TRADITIONAL TEACHING AND TESTING

There are several stories we like to tell to emphasize why we need substantial restructuring of the way assessment is done in schools. The first comes from Alan Schoenfeld (in press) who observed a geometry teacher in the Rochester New York schools who was reputed to be one of the best teachers in the state because his students did so well on the Regents exam in geometry. It turned out that he had his students memorize the twelve proofs that might be on the Regents exam, which is a complete perversion of the goal of learning geometry. A similar tale comes from Jerry Pines, whose son took an AP English course in which the students never wrote more than a one-page paper because that is the length of writing required for the AP exam.

Another story comes from Sig Abeles who, with Joan Baron, administered a test statewide in Connecticut at the eighth and twelfth grade levels on density (which is taught in the eighth grade). Students did quite well on a multiple-choice test item, where they were given the weight and volume and asked to figure out the density. But when they were given a block of wood, a ruler, and a scale, only about 3% of the eighth graders and 12% of the twelfth graders could solve the problem. Simply stated, students often learn to give back answers to written items that they have no ability to apply in real life.

The final story comes from Norman Frederiksen (1984), who during World War II was assigned to improve testing procedures for the job of gunner's mate in the Navy. This is a job that requires cleaning and maintaining guns on board ships, but he found that the teaching was by lecture and the testing was by paper and pencil. He proposed a performance test, based on the tasks that gunner's mates actually carry out. But the instructors objected to this because they thought the students would fail. And they did. Subsequently, teaching practice changed in the courses, so that fairly soon students learned to do just as well on performance tests as they had previously done on pencil-and-paper tests. A similar change is reported to have occurred when performance testing was introduced into the elementary school science curriculum in New York State. If we change the way we test students, it really does affect what is taught.

A SYSTEMS APPROACH TO ASSESSMENT

We have argued elsewhere (Frederiksen & Collins, 1989) that if we are going to have *systemically valid* tests (i.e., tests that foster the learning of the knowledge and skills that the test is designed to measure), then the tests must meet four criteria:

1. **Directness** refers to the degree that the test specifically measures the knowledge and skill we want students to achieve, as opposed to measuring indicator variables for that knowledge and skill. Often directness is sacrificed for the sake of "objectivity."

2. **Scope** refers to the degree to which all of the knowledge and skill required are assessed. If part is omitted, teachers and students will misdirect their teaching and learning in order to maximize scores on tests.

3. **Reliability** refers to the degree to which different judges assign the same score to an assessment. It is critical to achieve fairness in any assessment.

4. **Transparency** refers to the ability of those being assessed to understand the criteria on which they are being judged. If they are to improve their performance, the assessment must be transparent.

We would argue that if school assessment is going to meet the criteria of *directness* and *scope*, assessment must go beyond pencil-and-paper testing. Video and computer technologies provide very different media for recording student performances, and make it pos-

sible to construct assessments that more fairly represent the range of knowledge and skills toward which education should be directed.

Frederiksen and Collins (1989) also developed a set of principles for the design of systemically valid tests. Here we will briefly describe the components of such a testing system and the methods by which the system encourages learning. The components of the system are:

Set of tasks. The tasks should be authentic, ecologically valid tasks that are representative of the kinds of knowledge and skills expected of the students (Brown, Collins, & Duguid, 1989; Wiggins, 1989).

Criteria for each task and aspect of expertise. Performance on a task (or aspect of a task) should be evaluated in terms of a small number of criteria that the students understand. The criteria should be small in number so that students can focus on them, they should be learnable so that student efforts lead to improvement, and they should cover all aspects required for good performance in the task.

A library of exemplars. To insure reliability of scores and learnability, there needs to be a library of records of student performances. These exemplars should include critiques by master assessors in terms of the criteria. They should be available to everyone, particularly the testees.

A training system for scoring tests. There are three groups who must learn to reliably assess test performance: (a) master assessors, (b) coaches, who for students would be teachers, and (c) the testees. Master assessors are charged with maintaining standards, and must train teachers to coach students as to how to perform well.

The methods for fostering improvement on the test include:

Practice in self-assessment. Students should have practice evaluating their test performance, which is possible using recording technologies such as video or computers (Collins & Brown, 1988).

Repeated testing. Students should have opportunities to take the test multiple times so they can strive to improve their scores.

Feedback on test performance. When students take the test, there should be a review of their performance with a master assessor or coach to help them see how their performance might be improved.

Multiple levels of success. There should be various landmarks of success, so that students can strive to do better.

This briefly summarizes the design principles we proposed. They are elaborated in the Frederiksen and Collins (1989) paper.

THE ROLES OF DIFFERENT MEDIA IN ASSESSMENT

The three media—pencil and paper, computers, and video—provide three different views of students. Our goal in this section is to delineate some of the different abilities that each medium can tap in order to emphasize how to construct a broader view of students.

The strength of the computer is its ability to track the process of learning and thinking and to interact with students. This gives it a variety of ways to tap into aspects of students' abilities that the other media cannot:

1. **Computers can record how students learn with feedback.** Because it is possible to put students into novel learning environments where the feedback is systematically controlled by the computer, it is possible to assess how well or how fast different students learn in such environments (Collins, 1990a). This can provide a measure, not just of current performance levels, but of learning ability in a particular domain.

2. **Computers can record students' thinking.** Because computers can trace the process by which students maneuver through a problem or task, they can record various aspects of students' strategic processes (Collins, 1990a; Frederiksen & White, 1990). For example, it is possible to keep records of whether students systematically control variables when testing a hypothesis. It is also possible to look at their control or metacognitive strategies (Collins & Brown, 1988; Schoenfeld, 1985) to determine what they do when they are stuck, how long they pursue dead ends, etc. In summary, the ability to trace the problem-solving process gives computers a way to measure the strategic aspects of their knowledge.

3. **Computers can record students' abilities to deal with realistic situations.** Because computers can simulate real-world situations, like running a bank or

repairing broken equipment (Collins, 1990b), it is possible to measure students' abilities in understanding situations, integrating information from different sources, and reacting appropriately in real time. Paper and pencil and video really cannot simulate real situations, so only computers give us a view of people's practical intelligence; that is, their ability to deal with realistic situations.

Video provides a very different view of students' abilities because it can record their ongoing activities and explanations in rich detail. This makes it possible to evaluate other abilities:

1. **Video can record how students explain ideas and answer questions that challenge their understanding.** Oral presentation is critical to many aspects of life, and video enables us to capture student presentations in the same way we capture written presentations with paper and pencil. With video we can see how well students integrate words and diagrams as they explain things. It is also possible to see how they answer challenging questions that their audience poses, how they deal with counterexamples and counterarguments, and how they clarify points that are unclear to the audience.

2. **Video can record how well a student listens.** Because video is a richly detailed medium, it is possible to see how students listen to other students or adults, how well they ask questions, and critique or summarize what is said. Listening requires a variety of critical skills: communicating to the speaker what you don't understand, directing their discussion to the issues that are particularly important or relevant to your needs, elaborating or synthesizing their remarks. Video is the only medium that enables us to evaluate their listening ability.

3. **Video can record how well students cooperate in a joint task.** Because video can record students' interactions, it can be used to measure how well they work with their partners, offer constructive comments, and monitor their partners' understanding. The skills of cooperating are critical to almost every aspect of life, and yet they are discouraged in most current school practice.

4. **Video can record how students carry out tasks and perform experiments.** Because video can record students carrying out actions, it makes it possible to evaluate their ability to perform science experiments, use tools, follow instructions, or create new objects. That is to say, video gives us the ability to see how students are integrating their eyes, hands, voices, and minds.

Paper and pencil can provide a much broader view of students than is currently employed in most testing. The major uses of pencil and paper in current testing are to measure students' knowledge of facts, concepts, and procedures, their ability to solve problems, and their ability to comprehend text. Two additional ways that paper and pencil might profitably be used are:

1. **Paper and pencil can record how students compose texts and documents of different kinds.** Paper and pencil are sometimes used to evaluate how well students can write a persuasive essay, a clear explanation, or an interesting story, but it also should be used to evaluate students' reports, memos, letters, and even graphs, drawings, or musical scores. Much more sophisticated multimedia documents can be produced with computer tools, which may come to replace pencil and paper for document creation.

2. **Paper and pencil can record how students critique different documents or performances.** For example, students can be asked to critique the methodology of an experiment or the logic of an argument. They might be asked to review a play, concert, book, or dance performance. Students' critical abilities are rarely evaluated in current testing.

In this section we have tried to give an idea of the wide range of student abilities that are rarely, if ever, evaluated, and which the different media give us a means to document. Our argument is that current testing gives us a very narrow view of students, and this narrowness fundamentally misdirects all of education. It is critical that we extend the scope of testing to represent much more broadly the range of abilities necessary to being an educated person.

Many of the kinds of records proposed require subjective scoring, which some people object to as costly, time consuming, and inherently unfair. As we have argued elsewhere (Frederiksen & Collins, 1989), there are well-developed methods for achieving fairness in assessing student writing, and these methods are applicable to records from video and computers. Furthermore, the limits of what we know how to objectively score so fundamentally misdirect the educational enterprise that the real costs of objective scoring may far outweigh the costs of instituting a testing system that measures a broad range of student abilities.

TASKS EMPLOYING DIFFERENT MEDIA

We are currently trying to develop systemically valid methods of assessing student performance in the context of high school science. A key part of this work is to explore what kinds of tasks will enable students to use and demonstrate the broader range of abilities outlined above, and this requires very different kinds of tasks than are now the norm. Successful tasks are likely to have the following properties: they are complex enough to engage students in real thinking and performances; they exemplify "authentic" work in the disciplines; they are open-ended enough to encourage different approaches, but sufficiently constrained to permit reliable scoring; and appropriate records of student abilities can be readily collected and compiled for assessment purposes. We can illustrate the kinds of tasks that we are recommending, using computers and video, by describing some assessment tasks we have developed in the science project, and also some tasks developed by other researchers. For each task, we will also suggest different scoring criteria that might be employed for evaluating the student records.

One of the important issues in design of successful tasks concerns the kinds of records that are collected. These may take one or more forms, including the products of students' work; a finished presentation, performance, or verbal explanation; or aspects of students' thinking and problem-solving processes as they work on a task. Decisions about what process records to collect are interesting parts of our task development research. They might be "snapshots" of key parts of the task (e.g., what configuration of variables does a student select for a simulation). They might even be continuous recordings of students' reflections about their work. Essential to collecting records for assessment is that these records are efficient for scoring and that they capture the most important aspects of the different target abilities. It is also important that the collection of process records not have the undesirable systemic effect of constraining students' ways of working, so that they have to carry out tasks in a rigidly prescribed way.

Formulating relationships between variables. In our science project, we are collecting data using a computer program called Physics Explorer. Physics Explorer provides students with a simulation environment in which there are a variety of different models, each with a large set of associated variables that can be manipulated. Students conduct experiments to determine how different variables affect each other within a physical system. For example, one task duplicates Galileo's pendulum experiments, where the problem is to figure out what variables affect the period of motion. In a second task, the student must determine what variables affect the friction acting on a body moving through a liquid. Students might be evaluated in terms of the following traits: (1) how systematically they consider each possible independent variable; (2) whether they systematically control other variables while they test a hypothesis; (3) whether they can formulate qualitative relationships between the independent variables and the dependent variables; and (4) whether they can formulate quantitative relationships between the independent variables and the dependent variables.

Troubleshooting or diagnosing problems. Another kind of task that arises in many different settings is diagnosing why a system is not behaving as expected. Such problems are most common in computer programming, electronics, and medicine, but they can occur with any system, such as government or business. Using simulations of such systems, computers can provide students with a faulty version of a system, such as a circuit, and ask them to troubleshoot in order to find out why it is not doing what it is supposed to. Students' performances might be evaluated on such a task in terms of: (1) how they reason about a system's behavior in order to generate hypotheses about faults; (2) how systematically they collect data to evaluate their hypotheses; and (3) how consistent their hypothesis revisions are with the data they have collected.

Design. Computers provide a setting where students can carry out design tasks, such as designing a circuit, an ecosystem, or a governmental policy. The system can be tried out in a simulation, the effects of the design observed, and revisions made where appropriate. One possible task is for students to design a set of activities to teach younger students about Newton's Laws using a Dynaturtle in Logo (diSessa, 1984; White 1984). A Dynaturtle is moved by firing impulses, like a rocket in outer space, so that it makes it possible to see the behavior of an object in a frictionless environment. We might evaluate such a task in terms of: (1) how creative the design is; (2) how well the

students understand the subject matter; (3) how systematic or coherent the design is; (4) how well the design carries out its intended purpose; and (5) how polished the design is.

Learning with feedback. With many computer-simulation environments it is possible to give students feedback on what they have done and hints as to good strategies to use (Campione & Brown, 1990; Frederiksen & White, 1990). In such environments it is possible to evaluate students in terms of: (1) how much their performance improves during some fixed period; (2) how responsive they are to suggestions given them; (3) how much they rely on hints; and (4) their overall performance level on the task.

For video, students can be assessed in the following kinds of tasks:

Oral presentations. Students might be asked to present the results of their work on projects either to the teacher or the class as a whole. Such talks should include both a presentation portion, where clarification questions are permitted, and a questioning period, where the students are challenged to defend their beliefs. Students' presentations might be judged in terms of: (1) depth of understanding; 2) clarity; (3) coherence; (4) responsiveness to questions; and (5) monitoring of their listeners' understanding.

Paired explanations. This task makes it possible to evaluate students' ability to listen as well as to explain ideas. First, one student presents to another student an explanation of a project he or she has completed or a concept (e.g., gravity) he or she has been working on. Then the two students reverse roles. The students should use the blackboard or visual aids wherever appropriate. The explainers can be evaluated using the same criteria as for oral presentations. The listeners might be evaluated in terms of: (1) the quality of their questions: (2) their ability to summarize what the explainer has said; (3) their helpfulness in making the ideas clear; and (4) the appropriateness of their interruptions.

Joint problem solving. Another use of video is in judging students' ability to work together to solve problems. The joint problem-solving tasks can consist of hands-on science experiments, construction projects, textbook problems, etc. The criteria for evaluating student performance might change depending on the task, but could consist of the following kinds of characteristics: (1) helpfulness; (2) creativity; (3) understanding; (4) sharing of work; and (5) monitoring progress toward the goal.

The objective in developing tasks to assess student ability is to find tasks that represent the entire range of activities that are required in life. Because we have been concerned with assessing scientific ability, we have been trying to design tasks that address the full range of qualities it is important for scientists to develop. This leads to a very different kind of assessment than traditional science assessments, which test only for students' recall of facts, concepts, and procedures, and their ability to solve short, well-defined problems.

POSSIBLE OBJECTIONS TO SYSTEMICALLY VALID TESTING

There are a number of issues that critics raise about the kind of testing system we have proposed. These include the cost, the problem of cheating, and the dangers of using the system for surveillance, of teacher/parent prepping of students, and of exacerbating the difficulties of minorities in the school system.

With respect to the cost issue, it is certainly true that the kind of testing proposed is much more expensive to administer. We would argue that testing by an outside agency should be extremely limited in any case, and so the high costs might have an incidental benefit of reducing the amount of outside, "on-demand" testing in our schools. Ideally, much of students' in-class effort would go into producing products that they and their teachers try to evaluate. Some of those might eventually go into a portfolio that would be part of the submission to an outside testing agency. Costs can also be minimized by having trained teachers in each school conduct interviews with students that form part of the students' record to be evaluated by an outside agency. To reiterate, the real cost of the current testing system is its misdirection of education. Our view is that it should be possible to develop a cost-effective testing system that does not have perverse effects on education.

The problem of cheating can be serious in any portfolio testing scheme. The problem is less severe with video than with either written or computer records, since video documents real-time performance and it is difficult to falsify such a record. It is possible

to practice until the performance is quite smooth, but it should be possible for judges to evaluate spontaneity if such a characteristic is desired for certain records. However, the best way to deal with cheating on any portfolio submission is to conduct an interview with students about the portfolio in order to verify its authenticity. Such an interview can probe into different aspects of the portfolio, to determine how deeply the student understands the topics covered in the portfolio.

Some people worry that computers and videos will be used to maintain surveillance of students as part of their assessment function. For example, computer-based integrated learning systems that give students a sequence of tasks to work on, keep records of how each student does on each task and how they are progressing through the sequence. If a teacher is so inclined, it is possible to keep fairly close track of students with such a system. This type of surveillance raises issues of privacy and motivation: Will students come to feel that they are constantly being watched, and will they feel totally constrained to do everything according to the rules, allowing for no inventiveness or exploration? We do not think that this is the most effective use of computers for education (Collins, in press; Collins, Hawkins, & Carver, in press), but we think the best safeguard against such a danger is a portfolio system, where students decide what should be submitted for assessment.

The goal of the system is to encourage prepping of students by teachers and parents toward legitimate goals of education. Obviously parents or teachers, who care about education and who have the skills to do so, may coach students more than those who do not. This in turn could exacerbate the problems children from some minority cultures have, though not necessarily. If minority cultures value hands-on activity or oral language more than abstract thinking and written language (Gardner, 1990), the involvement of media that can capture different cultural emphases may offset coaching differences. As a society, we need to encourage all minority cultures to emphasize education for their children, and perhaps a testing system that provides them areas in which to excel will make this emphasis easier to realize.

There is the problem that many parents, including those from minority cultures, think that education must focus on the types of abilities currently embodied in tests. Our thesis is that there needs to be a fundamental change in public understanding of the goals of education. But such a change will only come very slowly, and it is likely to follow rather than precede any changes in the educational system (Collins, in press).

CONCLUSION

We are at the beginning of a program of research to demonstrate the reliability of an entirely new approach to assessment in schools. If it is viable, we would hope that it could be put in place in a number of schools and be used as an alternative form of testing for assigning student grades and admission to college. But the biggest challenges are still to come.

We would like to reiterate the problems that we see ahead (from Frederiksen & Collins, 1989). Clearly, much research needs to be done to test the assumptions on which our proposal is based. Can performances be reliably assessed on a common scale when the particular tasks that testees carry out may vary? Does an awareness of criteria help students to improve performance on projects and teachers to become more effective in the classroom? Can a consensus be reached on what are appropriate criteria for different domains and activities? Can scoring standards be met when assessment is decentralized? These and other questions are the focus of our research effort in support of a new, systemically valid system of educational testing.

References

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.

Campione, J. C., & Brown, A. L. (1990). Guided learning and transfer: Implications for approaches to assessment. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), *Diagnostic monitoring of skills and knowledge acquisition* (pp. 141-172). Hillsdale, NJ: Erlbaum.

Collins, A. (1990a). Reformulating testing to measure learning and thinking. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), *Diagnostic monitoring of skills and knowledge acquisition* (pp. 75-87). Hillsdale, NJ: Erlbaum.

Collins, A. (1990b). Cognitive apprenticeship and instructional technology. In L. Idol & B. F. Jones (Eds.), *Educational values and cognitive instruction: Implications for reform* (pp. 119-136). Hillsdale, NJ: Erlbaum.

Collins, A. (in press). The role of computer technology in restructuring schools. In K. Sheingold & M. Tucker (Eds.), *Restructuring for learning with technology*. Rochester, NY: Center for Education and the Economy.

Collins, A., & Brown, J. S. (1988). The computer as a tool for learning through reflection. In H. Mandl & A. Lesgold (Eds.), *Learning issues for intelligent tutoring systems* (pp. 1-18). New York: Springer-Verlag.

Collins, A., Hawkins, J., & Carver, S. M. (in press). A cognitive apprenticeship for disadvantaged students. In B. Means (Ed.), *Teaching advanced skills to disadvantaged students.*

diSessa, A. (1982). Unlearning Aristotelian physics: A study of knowledge-based learning. *Cognitive Science, 6*, 37-76.

Frederiksen, J. R., & White, B. Y. (1990). Intelligent tutors as intelligent testers. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 1-25). Hillsdale, NJ: Erlbaum.

Frederiksen, J. R. & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher, 18*(9), 27-32.

Frederiksen, N. (1984). The real test bias. *American Psychologist, 39*(3), 193-202.

Gardner, H. (1990). Assessment in context: The alternative to standardized testing. In B. Gifford & C. O'Connor (Eds.), *Future assessments: Changing views of aptitude, achievement, and instruction*. Boston: Kluwer.

Schoenfeld, A. H. (in press). On mathematics as sense-making: An informal attack on the unfortunate divorce of formal and informal mathematics. In D. N. Perkins, J. Segal, & J. Voss (Eds.), *Informal reasoning and education*. Hillsdale, NJ: Erlbaum.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, FL: Academic Press.

White, B. Y. (1984). Designing computer activities to help physics students understand Newton's laws of motion. *Cognition and Instruction, 1*, 69-108.

Wiggins, G. (1989, May). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 703-713.

Zuboff, S. (1988). *In the age of the smart machine: The future of work and power*. New York: Basic Books.